The Surprising Power of Belief Propagation

Elchanan Mossel

January 8, 2015

Why does BP Works??

- ML algorithm work well in practice.
- But Why?
- Understand linear algebra algorithms (spectral, singular values ...) well.
- Non-linear algorithms?
- Especially when they do better even in practice! than all algorithm we know.
- Today: Belief Propagation.

Graphical Models and Belief Propagation

- Graphical Models or Markov Random Fields are one of the most popular ways to prescribe high dimensional distributions.
- Encoding based on *conditional independence statements*.
- Based on a probabilistic model on graph / *graphical model*.

Graphical Models and Belief Propagation

- Graphical Models or Markov Random Fields are one of the most popular ways to prescribe high dimensional distributions.
- Encoding based on *conditional independence statements*.
- Based on a probabilistic model on graph / *graphical model*.
- (Pairwise) Graphical model is based on a graph $G = (V, E)$ and a distribution

$$p((x_v : v \in V)) = Z^{-1} \prod_{(u,v)\in E} \psi_{(u,v)}(x_u, x_v), \quad x \in A^V$$

Graphical Models and Belief Propagation

- Graphical Models or Markov Random Fields are one of the most popular ways to prescribe high dimensional distributions.
- Encoding based on *conditional independence statements*.
- Based on a probabilistic model on graph / *graphical model*.
- (Pairwise) Graphical model is based on a graph $G = (V, E)$ and a distribution

$$p((x_v : v \in V)) = Z^{-1} \prod_{(u,v) \in E} \psi_{(u,v)}(x_u, x_v), \quad x \in A^V$$

- Goal of Belief Propagation: Compute marginals:

$$p(x_v = a)??$$

Example: Graph coloring

- Let p be the uniform q-coloring model of a graph.

Example: Graph coloring

- Let p be the uniform q-coloring model of a graph.
- Can write:

$$p((x_v : v \in V)) = Z^{-1} \prod_{(u,v) \in E} \psi(x_u, x_v), \quad x \in [q]^V$$

where $\psi(c, d) = 1 - \delta_{c,d}$.

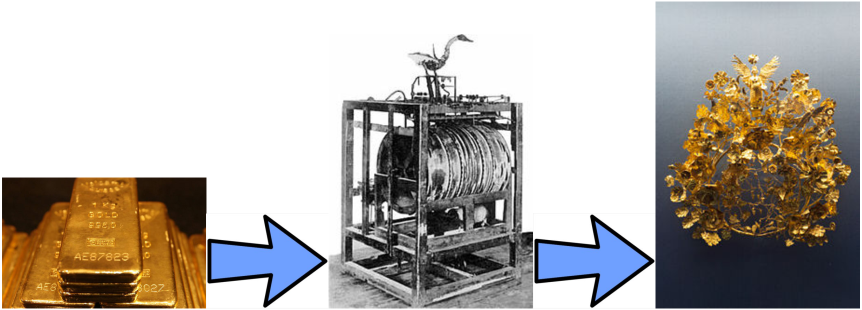
Example: Graph coloring

- Let p be the uniform q-coloring model of a graph.
- Can write:

$$p((x_v : v \in V)) = Z^{-1} \prod_{(u,v) \in E} \psi(x_u, x_v), \quad x \in [q]^V$$

  where $\psi(c, d) = 1 - \delta_{c,d}$.
- $$p(x_v = c)??$$

Belief Propagation on Trees



Belief Propagation on Trees

- On trees: $O(n^2)$ time to get all marginals using recursion.

Belief Propagation on Trees

- On trees: $O(n^2)$ time to get all marginals using recursion.
- More sophisticated *Dynamic Programing* is done in $O(n \times \text{diameter})$. "Belief Propagation".

Belief Propagation on Trees

- On trees: $O(n^2)$ time to get all marginals using recursion.
- More sophisticated *Dynamic Programing* is done in $O(n \times \text{diameter})$. "Belief Propagation".
- Belief Propagation Variables: $(\eta^a_{v \to u} : (v, u) \in E,\ a \in A)$.

Belief Propagation on Trees

- On trees: $O(n^2)$ time to get all marginals using recursion.
- More sophisticated *Dynamic Programing* is done in $O(n \times \text{diameter})$. "Belief Propagation".
- Belief Propagation Variables: $(\eta^a_{v \to u} : (v, u) \in E,\ a \in A)$.
- Updates:

$$\eta^a_{v \to u}(t+1) := Z^{-1} \prod_{w \neq u, (w,v) \in E} \sum_b \eta^b_{w \to v}(t) \psi_{(v,u)}(b, a)$$

Belief Propagation on Trees

- On trees: $O(n^2)$ time to get all marginals using recursion.
- More sophisticated *Dynamic Programing* is done in $O(n \times \text{diameter})$. "Belief Propagation".
- Belief Propagation Variables: $(\eta^a_{v \to u} : (v, u) \in E, \ a \in A)$.
- Updates:

$$\eta^a_{v \to u}(t+1) := Z^{-1} \prod_{w \neq u, (w,v) \in E} \sum_b \eta^b_{w \to v}(t) \psi_{(v,u)}(b, a)$$

- Marginal of x_u is approximated by

$$p(x_u = a) := Z^{-1} \prod_{(v,u) \in E} \eta^a_{v \to u}(\infty)$$

Belief Propagation on Trees

- On trees: $O(n^2)$ time to get all marginals using recursion.
- More sophisticated *Dynamic Programing* is done in $O(n \times \text{diameter})$. "Belief Propagation".
- Belief Propagation Variables: $(\eta^a_{v \to u} : (v, u) \in E,\ a \in A)$.
- Updates:

$$\eta^a_{v \to u}(t+1) := Z^{-1} \prod_{w \neq u, (w,v) \in E} \sum_b \eta^b_{w \to v}(t) \psi_{(v,u)}(b, a)$$

- Marginal of x_u is approximated by

$$p(x_u = a) := Z^{-1} \prod_{(v,u) \in E} \eta^a_{v \to u}(\infty)$$

- Example of 3-coloring:

$$\eta^a_{v \to w} = \frac{\prod_{u \in N(v) \setminus w} (1 - \eta^a_{u \to v})}{\sum_{b=1}^{3} \prod_{u \in N(v) \setminus w} (1 - \eta^b_{u \to v})}$$

Belief Propagation on (tree-like) graphs

- BP is very extensively applied to general graphs.

Belief Propagation on (tree-like) graphs

- BP is very extensively applied to general graphs.
- Not clear what it gives!

Belief Propagation on (tree-like) graphs

- BP is very extensively applied to general graphs.
- Not clear what it gives!
- Mathematical formulation:

Belief Propagation on (tree-like) graphs

- BP is very extensively applied to general graphs.
- Not clear what it gives!
- Mathematical formulation:
- Given a graph G, let $T(G)$ be the universal cover of G.
- $T(G)$ is the tree of non-backtracking walks on G.

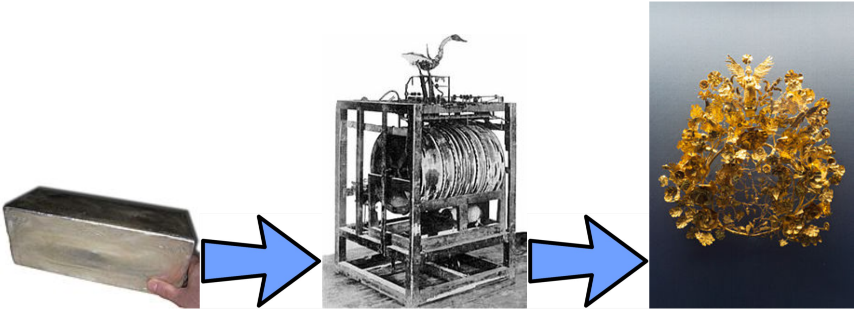
Belief Propagation on (tree-like) graphs

- BP is very extensively applied to general graphs.
- Not clear what it gives!
- Mathematical formulation:
- Given a graph G, let $T(G)$ be the universal cover of G.
- $T(G)$ is the tree of non-backtracking walks on G.
- To compute marginal x_v at G, compute x_v at $T(G)$.

Belief Propagation on (tree-like) graphs

- BP is very extensively applied to general graphs.
- Not clear what it gives!
- Mathematical formulation:
- Given a graph G, let $T(G)$ be the universal cover of G.
- $T(G)$ is the tree of non-backtracking walks on G.
- To compute marginal x_v at G, compute x_v at $T(G)$.
- If G is not a forest then $T(G)$ is infinite ...

BP on tree-like graphs and local information



Treelike graphs, local information and LDPC

- Beautiful Work in Coding Theory - LDPC

Treelike graphs, local information and LDPC

- Beautiful Work in Coding Theory - LDPC
- If $G = (V, E)$
  1. locally tree-like and
  2. can initialize $\eta_{u \to v}$ so that they are correlated to x_v

  Then BP converges to correct correct values!

Treelike graphs, local information and LDPC

- Beautiful Work in Coding Theory - LDPC
- If $G = (V, E)$
  1. locally tree-like and
  2. can initialize $\eta_{u \to v}$ so that they are correlated to x_v

  Then BP converges to correct correct values!
- Luby-Mitzenmacher-Shokrollahi-88
- Spielman-00,Richardson-Shokrollahi-Urbanke-01.

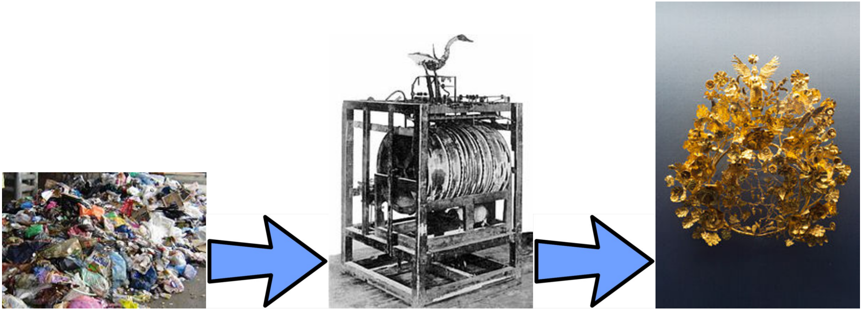
Treelike graphs, local information and LDPC

- Beautiful Work in Coding Theory - LDPC
- If $G = (V, E)$
  1. locally tree-like and
  2. can initialize $\eta_{u \to v}$ so that they are correlated to x_v

  Then BP converges to correct correct values!
- Luby-Mitzenmacher-Shokrollahi-88

- (Why) Does BP work in other cases?
- In particular, how does it work when there is no way to initialize the messages?

BP on tree-like graphs without local information



???

The Block Model

- Random graph $G = (V, E)$ on n nodes.
- Half blue / half red.

The Block Model

- Random graph $G = (V, E)$ on n nodes.
- Half blue / half red.
- Two nodes of the same color are connected with probability a/n.
- Two nodes with different colors are connected with probability b/n.

The Block Model

- Random graph $G = (V, E)$ on n nodes.
- Half blue / half red.
- Two nodes of the same color are connected with probability a/n.
- Two nodes with different colors are connected with probability b/n.
- Inference: find which nodes are red and which are blue ?

The Block Model

- Random graph $G = (V, E)$ on n nodes.
- Half blue / half red.
- Two nodes of the same color are connected with probability a/n.
- Two nodes with different colors are connected with probability b/n.
- Inference: find which nodes are red and which are blue ?
- Conjecture (Decelle, Krzakala, Moore and Zdeborova): "Belief-Propagation" is the optimal algorithm.
- and ... possible to do better than random iff $(a-b)^2 > 2(a+b)$.

The Block Model Conjecture

- Conjecture (Decelle, Krzakala, Moore and Zdeborova): "Belief-Propagation" is the optimal algorithm.

The Block Model Conjecture

- Conjecture (Decelle, Krzakala, Moore and Zdeborova): "Belief-Propagation" is the optimal algorithm.
- Works amazingly well both on real and simulated data.

The Block Model Conjecture

- Conjecture (Decelle, Krzakala, Moore and Zdeborova): "Belief-Propagation" is the optimal algorithm.
- Works amazingly well both on real and simulated data.
- Other algorithms we know do not work as well. In particular, completely fail when $(a-b)^2 \sim 2(a+b)$.

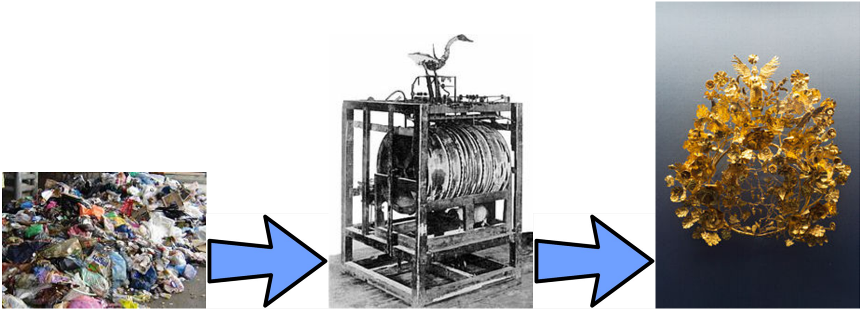
The Block Model Conjecture

- Conjecture (Decelle, Krzakala, Moore and Zdeborova): "Belief-Propagation" is the optimal algorithm.
- Works amazingly well both on real and simulated data.
- Other algorithms we know do not work as well. In particular, completely fail when $(a-b)^2 \sim 2(a+b)$.
- Note: can only solve up to global flip.

The Block Model Conjecture

- Conjecture (Decelle, Krzakala, Moore and Zdeborova): "Belief-Propagation" is the optimal algorithm.
- Works amazingly well both on real and simulated data.
- Other algorithms we know do not work as well. In particular, completely fail when $(a-b)^2 \sim 2(a+b)$.
- Note: can only solve up to global flip.
- Note: graph is very sparse - cannot hope to recover clusters exactly.

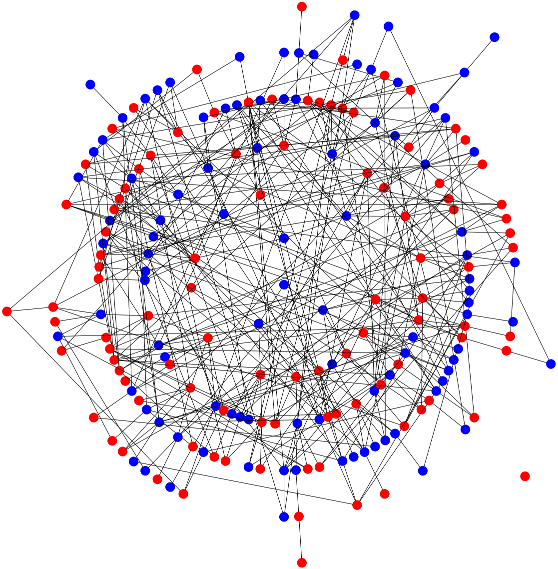
BP on tree-like graphs without local information



- Note: initializing correctly $(1/2, 1/2)$ is a fixed point.
- Instead initialize randomly ??

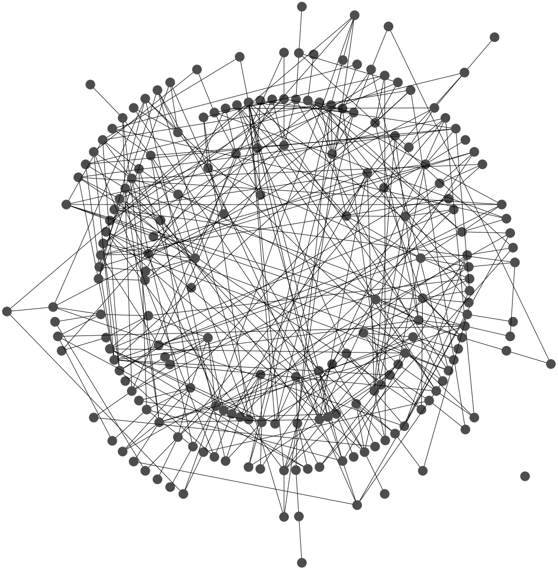
The Block Model in pictures

A sample from the model



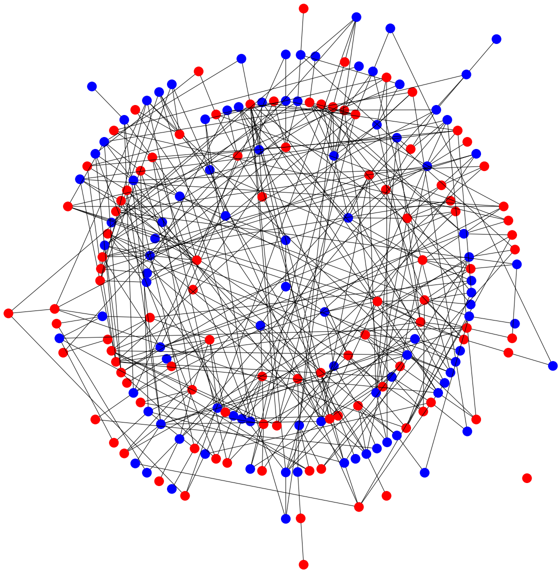
The Block Model in pictures

The data (one sample!)



The Block Model in pictures

What we want to Infer



The Conjecture is Correct

- **Thm 1** (M-Neeman-Sly 12): If $(a-b)^2 \leq 2(a+b)$ the impossible to infer better than random.

The Conjecture is Correct

- **Thm 1** (M-Neeman-Sly 12): If $(a-b)^2 \leq 2(a+b)$ the impossible to infer better than random.
- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect (infer better than random)

The Conjecture is Correct

- **Thm 1** (M-Neeman-Sly 12): If $(a-b)^2 \leq 2(a+b)$ the impossible to infer better than random.
- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect (infer better than random)
- **Thm 3** (M-Neeman-Sly, 14): If $(a-b)^2 > 100(a+b)$ then Belief Propagation is optimal for detection.

The Conjecture is Correct

- **Thm 1** (M-Neeman-Sly 12): If $(a-b)^2 \leq 2(a+b)$ the impossible to infer better than random.
- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect (infer better than random)
- **Thm 3** (M-Neeman-Sly, 14): If $(a-b)^2 > 100(a+b)$ then Belief Propagation is optimal for detection.
- **Thm 4** (M-Neeman-Sly, 14): It's possible to recover all the nodes in the graph **if and only if** w.h.p. *for all* nodes v, the majority of the neighbors of v are in the same cluster as v.

The Conjecture is Correct

- **Thm 1** (M-Neeman-Sly 12): If $(a-b)^2 \leq 2(a+b)$ the impossible to infer better than random.
- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect (infer better than random)
- **Thm 3** (M-Neeman-Sly, 14): If $(a-b)^2 > 100(a+b)$ then Belief Propagation is optimal for detection.
- **Thm 4** (M-Neeman-Sly, 14): It's possible to recover all the nodes in the graph **if and only if** w.h.p. *for all* nodes v, the majority of the neighbors of v are in the same cluster as v.
- Note: Thm 4 improves on a very long line of research in computer science and statistics including Boppana (87) Dyer and Freeze (89), Jerrum and Sorkin (89), Carson and Impagliazzo (01) and Condon and Karp (01).

BP and a New Type of Random Matrix

- **Thm 2** If $(a-b)^2 > 2(a+b)$ then possible to detect.

BP and a New Type of Random Matrix

- **Thm 2** If $(a-b)^2 > 2(a+b)$ then possible to detect.
- **Conj:**(Krzakala,Moore,M,Neeman,Sly, Zdebrovoa,Zhang 13): If A is the adjacency matrix, then w.h.p the second eigenvector of

$$N = \begin{pmatrix} 0 & D - I \\ -I & A \end{pmatrix}, \quad D = diag(d_{v_1}, \ldots, d_{v_n}),$$

is correlated with the partition.

BP and a New Type of Random Matrix

- **Thm 2** If $(a-b)^2 > 2(a+b)$ then possible to detect.
- **Conj:**(Krzakala,Moore,M,Neeman,Sly, Zdebrovoa,Zhang 13): If A is the adjacency matrix, then w.h.p the second eigenvector of

$$N = \begin{pmatrix} 0 & D-I \\ -I & A \end{pmatrix}, \quad D = diag(d_{v_1}, \ldots, d_{v_n}),$$

  is correlated with the partition.
- No orthogonal structure! N is not symmetric or normal. Singular vector of N are useless.

BP and a New Type of Random Matrix

- **Thm 2** If $(a-b)^2 > 2(a+b)$ then possible to detect.
- **Conj:**(Krzakala,Moore,M,Neeman,Sly, Zdebrovoa,Zhang 13): If A is the adjacency matrix, then w.h.p the second eigenvector of

$$N = \begin{pmatrix} 0 & D-I \\ -I & A \end{pmatrix}, \quad D = diag(d_{v_1}, \ldots, d_{v_n}),$$

  is correlated with the partition.
- No orthogonal structure! N is not symmetric or normal. Singular vector of N are useless.
- KMMNSZZ established connections between N and Belief Propagation

From BP to linear Algebra

- Coja-Oghlan-M-Vilenchik (09): to analyze BP with no local information linearize it.

From BP to linear Algebra

- Coja-Oghlan-M-Vilenchik (09): to analyze BP with no local information linearize it.
- Linearization gives $(n^2 - n) \times (n^2 - n)$ matrix.

From BP to linear Algebra

- Coja-Oghlan-M-Vilenchik (09): to analyze BP with no local information linearize it.
- Linearization gives $(n^2 - n) \times (n^2 - n)$ matrix.
- KMMNSZZ via Hashimoto 89 - get small matrix

$$N = \begin{pmatrix} 0 & D - I \\ -I & A \end{pmatrix}, \quad D = diag(d_{v_1}, \ldots, d_{v_n}),$$

From BP to linear Algebra

- Coja-Oghlan-M-Vilenchik (09): to analyze BP with no local information linearize it.
- Linearization gives $(n^2 - n) \times (n^2 - n)$ matrix.
- KMMNSZZ via Hashimoto 89 - get small matrix

$$N = \begin{pmatrix} 0 & D - I \\ -I & A \end{pmatrix}, \quad D = diag(d_{v_1}, \ldots, d_{v_n}),$$

- Study it and conjecture it's optimality.

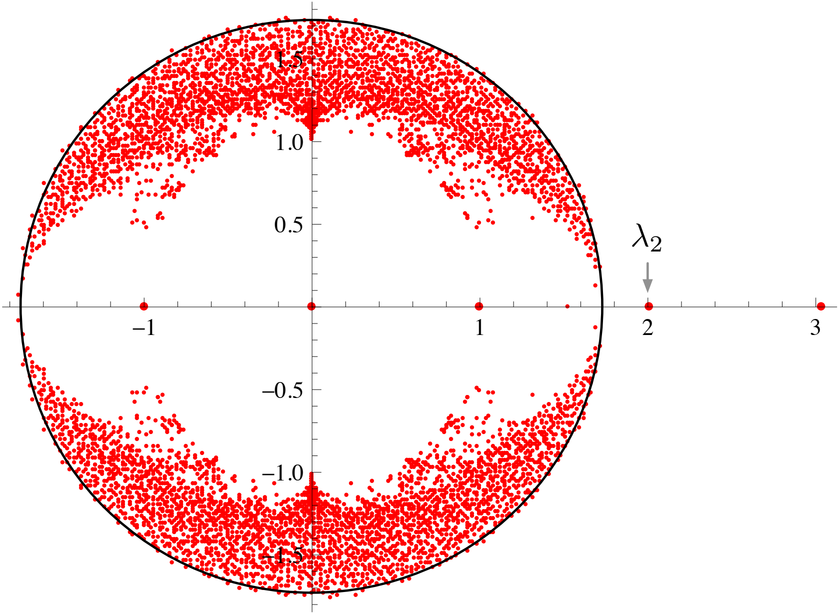
Zeta functions on graphs

1. Hashimoto-89: Introduced a graph analogue of Zeta functions of p-adic algebraic varieties:

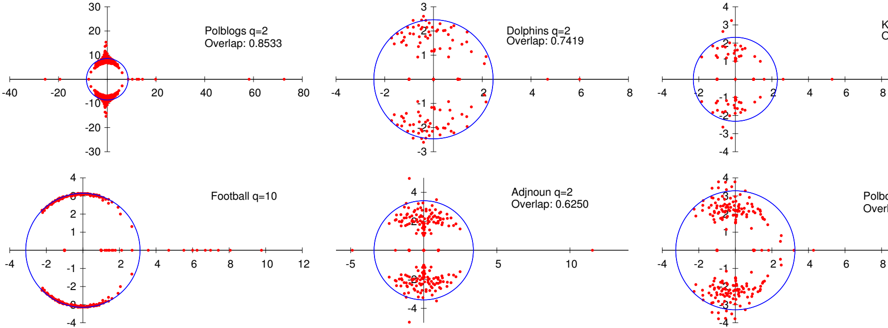
   $$Z(u,f) = \exp\Big(\sum_{\ell=1}^{\infty}\sum_{C\in X_\ell}\frac{f(C)}{\ell}u^\ell\Big),$$

   where $X_\ell =$ set of closed non backtracking loops of length ℓ and $f(C) = \prod_{e\in C} f(e)$.
2. Proved that $Z(f,u)$ is a rational function of u.
3. Asked: how much $Z(f,u)$ is revealing about the graph ...

The Spectrum of N



The spectrum on real networks



Performance on Real Networks

- $R = N$.
- $L =$ normalized laplacian (random walk matrix).

network name	BP overlap	sign of vector 2 of $\mathbf{R}$	k-means of $\mathbf{R}$	sign of vector 2 of $\mathbf{L}_{sym}$	k-means of $\mathbf{L}_{sym}$
words	*	**0.9107**	0.875	0.5625	0.5714
political blogs	0.5167	0.9313	0.6383	**0.9542**	0.9476
karate club	0.5588	**1**	**1**	0.9706	**1**
dophin	**0.9838**	0.8710	0.96774	0.9677	**0.9839**
brsmall	*	0.6548	**0.69345**	0.6235	0.6687
brcorp	*	0.6993	0.72631	**0.7332**	0.6993
adjnoun	0.5625	0.8125	**0.8214**	0.5446	0.5357

Two proofs avoiding the spectral conjecture

- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect.

Two proofs avoiding the spectral conjecture

- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect.
- MNS: Let $X^{\ell}(u,v) = \sum_{\Gamma} \prod_{e \in G}(1((u,v) \in G) - \frac{a+b}{2})$ where the sum is over all non backtracking walks of length $\ell = C \log n$.

Two proofs avoiding the spectral conjecture

- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect.
- MNS: Let $X^\ell(u,v) = \sum_\Gamma \prod_{e \in G}(1((u,v) \in G) - \frac{a+b}{2})$ where the sum is over all non backtracking walks of length $\ell = C \log n$.
- Show that $X^\ell(u,v)$ is (typically) larger if u and v are in same cluster.

Two proofs avoiding the spectral conjecture

- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect.
- MNS: Let $X^{\ell}(u,v) = \sum_{\Gamma} \prod_{e \in G}(1((u,v) \in G) - \frac{a+b}{2})$ where the sum is over all non backtracking walks of length $\ell = C \log n$.
- Show that $X^{\ell}(u,v)$ is (typically) larger if u and v are in same cluster.
- Massoulie: Define a symmetric matrix $A_{u,v} =$ number of self-avoiding walks from u to v of length $\varepsilon \log n$ and show second eigenvector is correlated with partition.

Two proofs avoiding the spectral conjecture

- **Thm 2** (M-Neeman-Sly, Massoulie 14): If $(a-b)^2 > 2(a+b)$ then possible to detect.
- MNS: Let $X^\ell(u,v) = \sum_\Gamma \prod_{e \in G}(1((u,v) \in G) - \frac{a+b}{2})$ where the sum is over all non backtracking walks of length $\ell = C \log n$.
- Show that $X^\ell(u,v)$ is (typically) larger if u and v are in same cluster.
- Massoulie: Define a symmetric matrix $A_{u,v} =$ number of self-avoiding walks from u to v of length $\varepsilon \log n$ and show second eigenvector is correlated with partition.
- Massoulie gets symmetric matrix. MNS - almost linear time.

Future Research

- Other planted models: more than two clusters, unequal size etc.

Future Research

- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.

Future Research

- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.
- For example: hidden clique.

Future Research

- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.
- For example: hidden clique.
- More challenging: BP and Survey Propagation for satisfiability problems.

Future Research

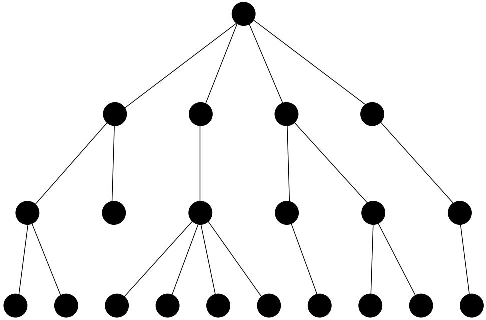
- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.
- For example: hidden clique.
- More challenging: BP and Survey Propagation for satisfiability problems.
- How to let linear algebra algorithms utilize local information?

A New Type of Phase Transition Question

- **Thm 1** (M-Neeman-Sly 12): f $(a-b)^2 \leq 2(a+b)$ the impossible to infer better than random.
- **Thm 3** (M-Neeman-Sly, 14): If $(a-b)^2 > 100(a+b)$ then Belief Propagation is optimal for detection.
- Proofs via phase transitions for broadcasting on trees. Thm 3 requires a new phase transition!

Broadcasting on trees

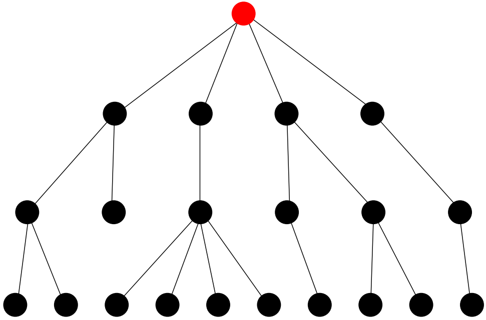
Take a tree. Fix $\epsilon \in (0, 1)$.



Broadcasting on trees

Take a tree. Fix $\epsilon \in (0, 1)$.

Color the root randomly.

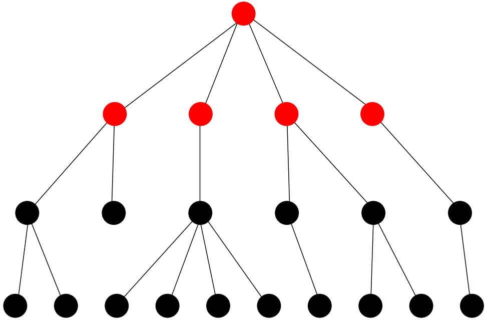


Broadcasting on trees

Take a tree. Fix $\epsilon \in (0, 1)$.

Color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color



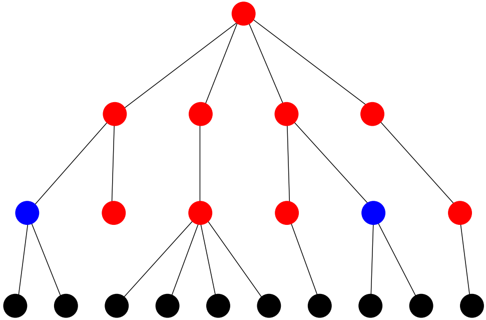
.

Broadcasting on trees

Take a tree. Fix $\epsilon \in (0, 1)$.

Color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color



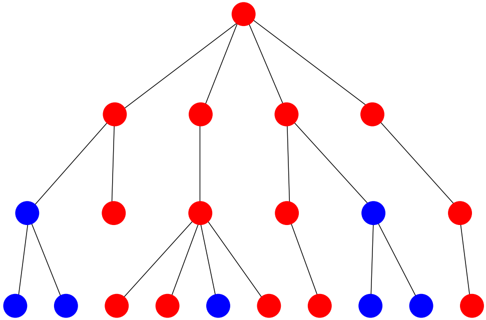
.

Broadcasting on trees

Take a tree. Fix $\epsilon \in (0, 1)$.

Color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color



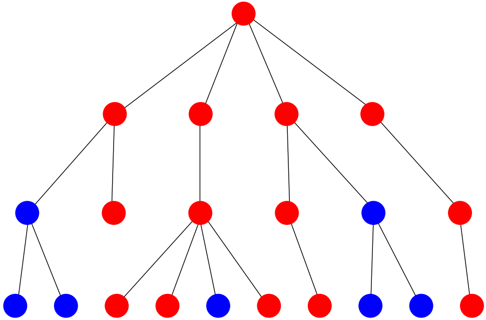
.

Broadcasting on trees

Take a tree. Fix $\epsilon \in (0, 1)$.

Color the root randomly.

For each child, copy the color
with probability $1 - \epsilon$.
Otherwise, flip the color



Question: given the leaves, can we guess the color of the root?

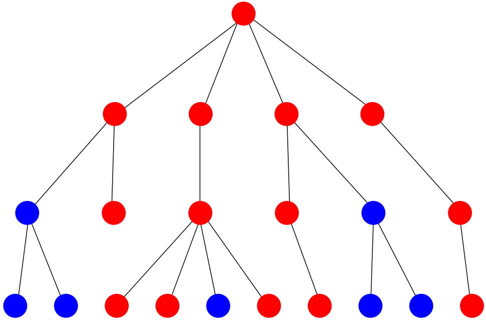
.

Broadcasting on trees

Take a tree. Fix $\epsilon \in (0, 1)$.

Color the root randomly.

For each child, copy the color
with probability $1 - \epsilon$.
Otherwise, flip the color



Question: given the leaves, can we guess the color of the root?
Answer: iff $(1 - 2\epsilon)^2 d > 1$
(where d is the *branching number* of the tree)
(... Evans, Kenyon, Peres, Schulman, 2000 ...)

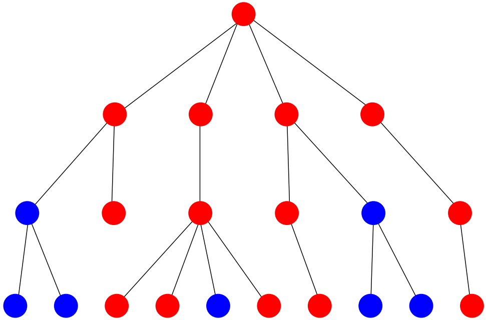
.

Broadcasting on trees

Take a tree. Fix $\epsilon \in (0, 1)$.

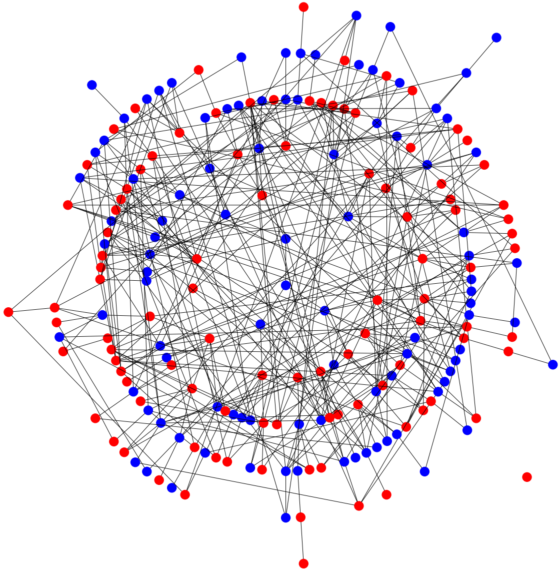
Color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color

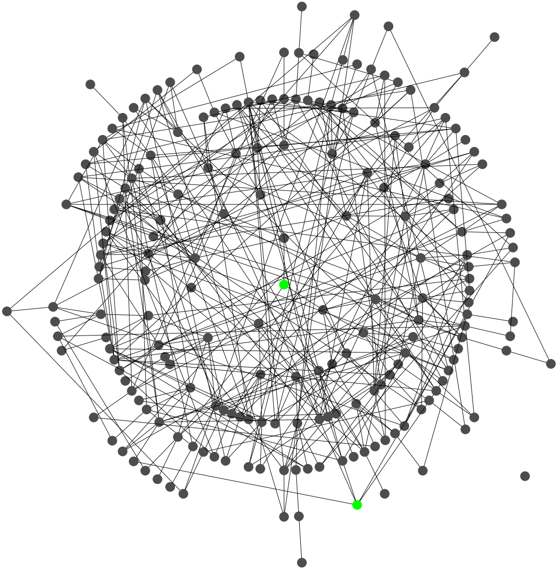


Question: given the leaves, can we guess the color of the root?
Answer: iff $(1 - 2\epsilon)^2 d > 1$
(where d is the *branching number* of the tree)
(... Evans, Kenyon, Peres, Schulman, 2000 ...)
$\implies$ Thm 1 with $\varepsilon = a/(a + b), d = (a + b)/2$.

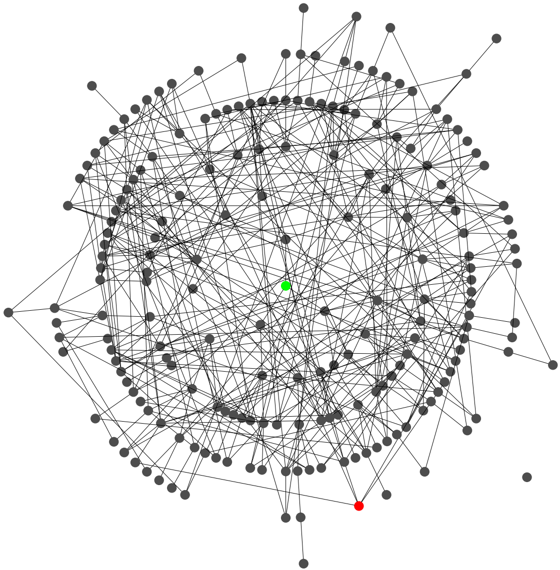
Back to the original problem



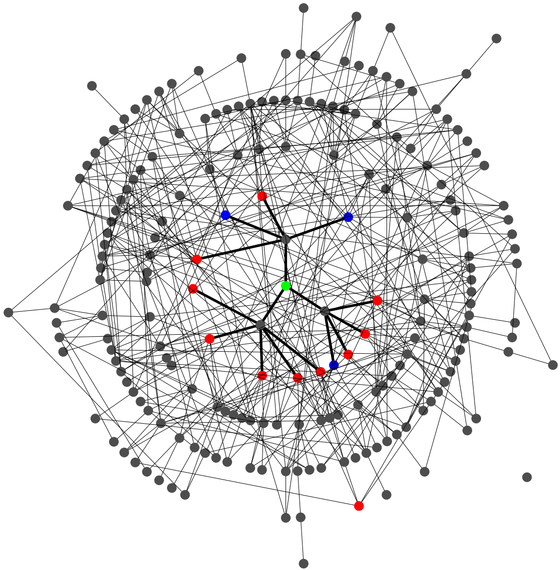
Back to the original problem



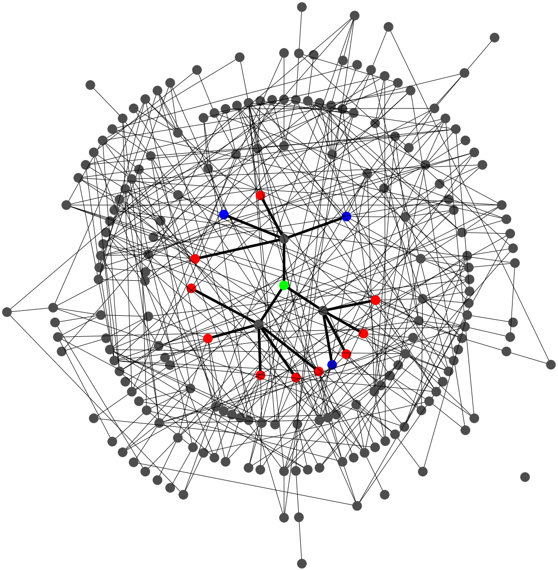
Back to the original problem



Back to the original problem



Back to the original problem

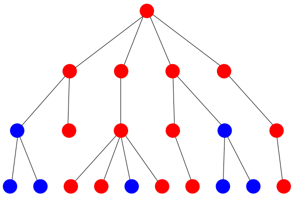


Robust tree reconstruction

To Analyze BP with good initial messages, we need to understand the following process

Robust tree reconstruction

To Analyze BP with good initial messages, we need to understand the following process



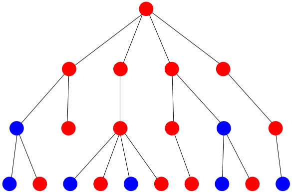
Take a tree and color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color

.

Robust tree reconstruction

To Analyze BP with good initial messages, we need to understand the following process



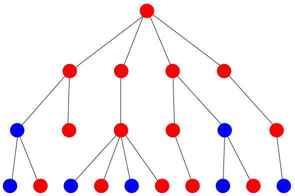
Take a tree and color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color

Flip the leaves with probability $\delta < 1/2$

Robust tree reconstruction

To Analyze BP with good initial messages, we need to understand the following process



Take a tree and color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color

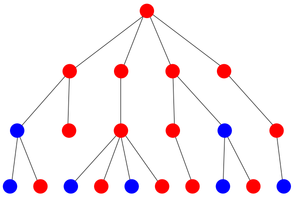
Flip the leaves with probability $\delta < 1/2$

Theorem (MNS-14)

If $(1 - 2\epsilon)^2 d \geq C$ then as $n \to \infty$, the extra noise doesn't hurt the reconstruction probability.

Robust tree reconstruction

To Analyze BP with good initial messages, we need to understand the following process



Take a tree and color the root randomly.

For each child, copy the color with probability $1 - \epsilon$. Otherwise, flip the color

Flip the leaves with probability $\delta < 1/2$

Theorem (MNS-14)

If $(1 - 2\epsilon)^2 d \geq C$ then as $n \to \infty$, the extra noise doesn't hurt the reconstruction probability.

Strong property of a non-linear dynamical system (stronger than non-ergodicity, "robust reconstruction" etc. (Janson-M-04).

Summary

Takeaways:

- Know how to reconstruct block models

Summary

Takeaways:

- Know how to reconstruct block models
- Theory can learn from practice.

Summary

Takeaways:

- Know how to reconstruct block models
- Theory can learn from practice.
- Nice to work with physicists and/or statisticians.

Future Research

- Other planted models: more than two clusters, unequal size etc.

Future Research

- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.

Future Research

- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.
- For example: hidden clique.

Future Research

- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.
- For example: hidden clique.
- More challenging: BP and Survey Propagation for satisfiability problems.

Future Research

- Other planted models: more than two clusters, unequal size etc.
- Typically expect computational threshold to be different than information threshold.
- For example: hidden clique.
- More challenging: BP and Survey Propagation for satisfiability problems.
- How to let linear algebra algorithms utilize local information?